**Vamsi Karumanchi**

Email: vamc317@gmail.com

PH: + 1 (571)-399-3055 **Data Engineer**

**Professional Summary**

* 7+ years of dedicated expertise in designing, developing, and implementing efficient data models and Business Intelligence solutions.
* Profound expertise in architecting and optimizing intricate data models tailored to diverse business needs.
* Ensured data models are scalable, adaptable, and performance-optimized to handle large datasets efficiently.
* Developed comprehensive Business Intelligence solutions enabling data visualization, analytics, and actionable insights
* Designed intuitive dashboards facilitating real-time insights for stakeholders across various domains.
* Demonstrated proficiency in utilizing a wide array of technologies, including Big Data tools (Hive, HDFS, Spark), cloud services (AWS, Azure), and programming languages (Python, Java, Scala).
* Skilled in executing ETL processes using tools such as Informatica and Apache Airflow.
* Translated complex business requirements into tangible and actionable insights, empowering data-driven decision-making.
* Enabled organizations to leverage data for strategic planning and achieve key business objectives.
* Successfully led the implementation of intricate data models, optimizing operations and enhancing overall efficiency within organizations.
* Designed and deployed intuitive BI dashboards providing real-time insights, enabling proactive decision-making, and improving business outcomes.
* Led cross-functional teams in seamlessly integrating Big Data technologies, resulting in scalable, agile, and business-centric data solutions.
* Demonstrated effective collaboration with diverse teams and stakeholders, ensuring alignment between technical solutions and overarching business objectives.
* Communicated technical complexities in easily understandable terms to non-technical stakeholders, fostering a shared understanding of data-driven strategies.

**Technical Skills:**

**Languages**: Python, SQL, T-SQL, PL/SQL

**Databases**: MySQL, MS-SQL Server, Oracle, Teradata, Snowflake.

**NOSQL** **Databases**: Dynamo DB, HBASE.

**Operating systems:** Windows, Linux

**ETL TOOLS:** SSIS, BODS, Azure Data Factory(V2)

**Data** **Visualization** **Tolls**: Tableau, PowerBI,SSRS.

**Cloud** **Services**: AWS (EC2, S3, EMR, RDS, Lambda, Redshift, Cloud Formation, Glue etc.), Azure Databricks, Azure Data Factory, Azure SQL, Snowflake, GCP (Data Proc, Composer, Big Query, Storage Bucket)

**PROFESSIONAL EXPERIENCE**:

**Anthem, Inc, Norfolk, VA Feb2022-Till date**

**Data Engineer**

**Responsibilities:**

* Implemented best practices for designing models supporting Dimensional and Relational data structures, optimizing database performance in collaboration with **DBAs** and **developers**.
* Collaborated closely with BI and Analytics teams, translating business requirements into effective data models to enhance **analytical** and **reporting capabilities**.
* Created advanced conceptual, logical, and physical data models using **ERWin**.
* Created advanced **ERWin** data models supporting both Dimensional and Relational structures for a major client in the real estate sector.
* Collaborated with cross-functional teams to analyze legacy systems and design data flow interfaces.
* Explored and incorporated big data technologies and **NoSQL** databases into data modeling strategies.
* Implemented best practices for designing models supporting Dimensional and Relational data structures, optimizing database performance in collaboration with DBAs and developers.
* Utilized **SQL** for data analysis and designing **STTM** (Source to Target Mapping) creations.
* Translated intricate business requirements into effective data models catering to analytical and reporting needs.
* Kept up to date with industry trends, ensuring data models aligned with evolving standards and methodologies.
* Collaborated with technical teams to optimize database performance, enhancing query efficiency by 25% within a six-month timeframe.
* Conducted performance tuning activities to enhance query performance and overall system efficiency.
* Created and managed comprehensive documentation, including entity-relationship diagrams and metadata.
* Worked directly with the client's **BI** and Analytics teams, translating complex business requirements into efficient **data models**, enabling enhanced analytical insights.
* Conducted detailed quality assurance checks, reducing data anomalies by 30% through **meticulous** model **inspection** and **rectification**.
* Created detailed documentation comprising entity-relationship diagrams and **metadata**, facilitating seamless project handovers and maintenance.

**Environment**: AWS Services, Hadoop (HDFS, Map Reduce), Python, Snowflake, Scala, IAM, PostgreSQL, Spark, Impala, Hive, Mongo DB, DevOps, HBase, Oracle, NIFI, Git, AWS Services (Lambda, EMR, Auto scaling).

**Infosys-Hyderabad, INDIA Mar2020-Nov2021**

**Walmart Inc, Bentonville, AR (Client USA)**

**Data Engineer**

**Responsibilities:**

* Collaborated within a multi-clustered Hadoop ecosystem, focusing on **Business Intelligence (BI)** solutions for diverse healthcare clients utilizing multiple clusters.
* Reviewed and managed log files using **HBase**, ensuring streamlined data retrieval and analysis for informed BI decisions.
* Contributed to the design and implementation of ingest architecture, leveraging Collibra workflow for process development and documentation to support business needs and test case capabilities.
* Contributed to optimizing MapReduce programs for improved data processing efficiency, employing strategies such as combiners, partitioners, and custom counters.
* Played a key role in data extraction and transformation from Hadoop databases using **Scoop**, collaborating closely with Linux systems and **Hadoop** databases for effective data operations.
* Managed the staging of real-time data from gateways into cloud storage systems (**AWS S3, Azure Blob, Google Cloud Storage**), preparing raw data for comprehensive BI analysis and reporting.
* Leveraged BigQuery extensively for in-depth data analysis, model building, and creating data flow pipelines, showcasing expertise in all SDLC phases.
* Implemented **ETL** pipelines using **Azure Data Factory**, **Python**, and **Snowflake** **SQL**, ensuring smooth data transformation and processing within and outside data warehouses.
* Utilized diverse technologies including **Pyspark, SQL, Python, Java, JSON** and **YAML** for efficient data ingestion and transformation tailored for multiple clients.
* Leveraged Apache Databricks (**Databricks**) for data analysis, measure calculations, and data visualization. Contributed to logic and mapping in Confluence for improved accessibility for developers.
* Actively involved in designing and executing **HIVE** queries and functions for data evaluation, filtering, loading, and storage, ensuring optimized BI data handling.
* Contributed to the implementation of data ingestion pipelines using **Apache** **Kafka**, **Flume**, and **Spark**, **storing**, and processing BI-relevant data efficiently into **HDFS**.
* Proficiently worked between **GCP** and **Azure** **clouds**, using **Azure** Data Factory for seamless data movement and management across platforms.
* Contributed to developing pipelines for incremental loads, facilitating operational teams to efficiently run and manage these pipelines in a production environment.
* Played an integral role in HEDIS sample and test deck analysis, collaborating on building the HEDIS parser for deck loading and synchronization with client information for score key calculations.

**Environment:** Hadoop, Python, Spark, Snowflake, Spark SQL, Test-Driven Development (TDD), Spark-Streaming, NoSQL, HBase, Data Lake, Zookeeper, Azure, Databricks, Azure Data Factory (ADF), Unix/Linux Shell Scripting, PyCharm, Informatica PowerCenter, Linux Shell Scripting, Airflow.

**Straive – USA Client April 2019- Feb2020**

**SQL Developer**

**Responsibilities:**

* Worked for ZILLOW (USA)client.
* Download the entire Zip file county from the ExaVault.
* Loading the County Source files in **SQL Server** input tables. Past Data: Migrate (i.e., Restore) the **PostgreSQL** County DB into **SQL** Server
* We Follow the "Zillow Guidelines " and county specific mapping to translate the input data to output tables.
* For each table we can create separate stored procedure to load data into the table.
* We can run General Zillow **QC procedure** for to find data issue and clear those issues.
* We can create new DB in **PostgreSQL** then load data into **PostgreSQL** tables.
* After load data into **PostgreSQL** tables, we can generate backup **schema** and validate that backup file.
* If we can get any Validation errors, we can do changes and clear those, Errors.
* After validation We can load that backup test **schema** file in **Exavault**.
* Once Client Approved Test Schema Then We can upload that Test schema File in Production on Exavault.

**Environment**: SQL Server, SSIS, SSRS, PgAdmin, Microsoft Excel, RDBMS, Power BI

**NOVAC TECHNOLOGY SOLUTIONS - Hyderabad, INDIA Dec 2017– Mar 2019**

**Software Engineer**

**Responsibilities:**

* Worked in Shriram Life Insurance(USA) client.
* Worked in creating **SQL** objects like views and procedures.
* Worked with different types of joins in **SQL Server** to get data from multiple tables based on the requirement.
* Creating Tables, stored procedures as per requirement.
* Worked with different types of **SQL functions** (String, Aggregate functions.)
* Working with Lookup functions in designer and scheduling the **batch jobs** using Management console.
* Expertise in managing insurance-related data, including policy data, customer information, claims data, and financial records. Ability to design and maintain **databases** that efficiently store and retrieve this critical information.
* Experience in extracting data from various sources, such as insurance databases, spreadsheets, and external data feeds. Proficiency in transforming and cleaning data to ensure its accuracy and consistency.
* Optimized **SQL** queries for retrieving and analyzing insurance data. Ability to tune **queries** for performance and efficiency, especially when dealing with large datasets.
* Collaborated with cross-functional teams, including business analysts, actuaries, underwriters, and IT professionals, to understand and address data-related requirements.
* Creating compelling and insightful data visualizations using **Power BI, Tableau**. Transform complex insurance data into easy-to-understand dashboards, reports, and charts.
* Integrating data from various sources, such as insurance databases, claims systems, policy management systems, and external sources, into **Power BI, Tableau** for comprehensive analysis.
* Using **Power BI, Tableau** to analyze insurance claims data, premium and loss analysis. Experience in identifying fraud patterns, claims trends, and outliers to help insurance companies make informed decisions.

**Environment**: SQL Server, SSIS, SSRS, PgAdmin, Microsoft Excel, RDBMS, R, Microsoft power BI, Tableau, Data Visualization tools, Microsoft Excel, VBA Macros.